Name	Arguments	Dependencies	Summary	Caller-callee
Sampling				
gene_locator.sh	<Gene list file>	None	Takes a list of genes and an annotation file (GFF), generates a file with the genomic locations of the genes (add output file format)	None
sampler.py	<script.py input_txt country_name num_samples output_directory>	None	This script samples a given number of samples ID	called by multysampler.py
multysampler.py	<country_list_file> <num_samples> <samplesID_file> <samples_years_file> <output_directory>	sampler.py	Calls the sampler function for a given list of countries	calls sampler.py
vcf_by_sample.py	<input_vcf_file> <sample_list_file> <output_file>	BCFtools	THIS SCRIPT TAKES A VCF FILE, A SAMPLE LIST FILE AND AN OUTPUT FILE NAME AS INPUTS, IT WILL RETURN A VCF FILE CONTAINING ONLY THE INFORMATION FROM THE SPECIFIED SAMPLES	called by vcf_cutter.sh calls *BCFtools*
vcf_cutter.sh	<gene_locations_file> <sample_list_directory>	vcf_by_sample.py BCFtools	This script takes a list of genes, and a directory containing a sample list for each country and it will cut a vcf file into individual vcf files for each gene for each country including only the samples in the list.	calls vcf_by_sample.py and *BCFtools*
Dxy				

Dxy.py	<population1.vcf> <population2.vcf> <genomic_region_start> <genomic_region_end>	Numpy Scikit-allel	This script calculates the average pairwise nucleotide diversity (Dxy) between 2 populations." + "\n\n" + "Usage: python ./Dxy.py population1.vcf population2genomic_region_end	called by dxy.sh
dxy.sh	<pop1_files.txt> <pop2_files.txt> <output_file> <gene_start> <gene_end>	Dxy.py	this will calculate pairwise dxy for every possible pair of populations given two directories each containing several population-level VCF files	calls Dxy.py
dN/dS				
BED_extractor.sh	<gene_ID> <gff_file> <output_dir>	None	This script will create a BED file with the genomic coordinates of introns and exons of a gene	
vcf_to_DNA.sh	<input.vcf> <genomic_region>	BCFtools SAMtools remove_indels.py	This script generates a DNA sequence for each sample in the VCF file. It calls BCFtools consensus for each sample, it will mask all the insertions using low caps and the insertions using *. Finally, it concatenates all the sequences into a single fasta file	called by vcf_to_cds.sh calls *SAMtools* and *BCFtools* calls remove_indels.py
CDS.py	fasta_file bed_file output_file	BioPython	This script will take a DNA sequence, its genomic location, and a reference file and output the CDS for the sequence.	called by vcf_to_cds.sh

vcf_to_cds.sh	<vcf_dir> <genomic_region> <bed_file>	calls vcf_to_DNA.sh vcf_to_cds	This script will take as an argument a directory containing several VCF files from the same gene but different populations, the genomic region, and a gene bed file. it will call the vcf_to_DNA.sh and CDS.py scripts to produce a fasta file containing the CDS for each population VCF file, the files will be saved to a folder called sequences inside the VCF files directory.	calls vcf_to_DNA.sh and vcf_to_cds
fasta_to_phylip_sequential .py	input.fasta output.phy	BioPython	This script takes a fasta file counting CDS sequences of equal length and converts them into Philip sequential format to input them into PAML-YN00	called by dnds.sh
dnds.sh	pop1_fasta_files_dir pop2_fast_files_dir output_file	PAML-YN00 fasta_to_phylip_seq uential.py	This script takes to directories containing several population-level CDS fasta files and will calculate pairwise dnds for each possible pair of populations	calls fasta_to_phylip_sequential.p y and *PAML-YN00*
Ploting				
heatmap.py	dnds_file Gene_name	Seaborn Matplot	Generates heatmaps from dn, ds, and dnds this needs to be modified so it can work with any countries	
heatmap2.py	dxy_file Gene_name	Seaborn Matplot	Generates heatmaps for Dxy this needs to be modify so it can work with any countries	